Electronic Supplementary Information:

Towards the role of metal ions in the structural variability of proteins: Cd^{II} speciation of a metal ion binding loop motif

Attila Jancsó, Dániel Szunyogh, Flemming H. Larsen, Peter W. Thulstrup, Niels J. Christensen, Béla Gyurcsik, Lars Hemmingsen**

Synthesis of Ac-SCHGDQGSDCSI-NH_2 (**HS**):

The peptide was prepared by solid phase peptide synthesis using the Fmoc methodology (Fmoc = 9-fluorenylmethoxycarbonyl). Rink Amide AM resin (Novabiochem, 200-400mesh, loading: 0.68 mmol/g) was used as a solid support. The amino acid building blocks were applied in 4-fold excess over the capacity of the resin. The amino acid residues were coupled to each other (and to the resin) by applying HBTU (3.8 eq./building block), HOBt (4 eq./building block) and N,N-diisopropylethylamine (7.8 eq./building block) in NMP. The Fmoc-protecting groups were removed by using a solution of 20% piperidine in NMP. The usual coupling reaction time was 1 h. The attachment of each amino acid residues was monitored by Kaiser-test (E. Kaiser, R.L. Colescott, C.D. Bossinger, P.I. Cook, *Anal. Biochem*., 1970, **34**, 595-598.), and by the detection of the cleaved Fmoc group at 304 nm in DMF. In the case of successful coupling, the free residual amino nitrogens were acetylated with the mixture of acetic anhydride, dichloromethane and pyridine (10–85–5%). After the last coupling step the Fmoc group was cleaved from the amino group of the N-terminal amino acid (serine) which was acetylated afterwards. Following acetylation the resin was rinsed by dichloromethane and methanol and then dried in vacuum. Cleavage of the peptide from the resin was performed in the mixture of TFA, H_2O, 1,2-Ethanedithiol and triisopropylsilane (94–2,5–2,5–1 %). The peptide was preciptated in cold diethyl ether, dissolved in water and purified by RP-HPLC using a Supelco Discovery BIO Wide Pore C18 (25×10 mm, 5 μm) semi preparative column. The dodecapeptide was eluted by using the mixtures of water and acetonitrile containing 0.05% TFA (Eluent A: H_2O-CH_3CN 93:7 %v/v with 0.05% TFA, Eluent B: H_2O-CH_3CN 7:93 %v/v with 0.05% TFA) applying the following program:

0-73 min: 0% B (isocratic); 73-83 min: 0-50% B (linear gradient); 83-86 min: 50% B (isocratic); 86-89 min: 50-0% B (linear gradient); 89-95 min: 0% B (isocratic).

Table S1: ^{1}H NMR (500 MHz) chemical shifts (ranges) (δ/ ppm) of the various proton resonances of **HS** (pH = 5.37, T = 298 K, H_2O/D_2O = 90:10 %v/v)

Residue	NH	$C_\alpha H$ [a]	$C_\beta H_2$ [b]	Others
Ser-1	8.34	4.45-4.52	3.83-3.97 [c]	
Cys-2	8.46	4.53-4.59	2.94	
His-3		~4.77	3.22 and 3.35	$C_{\varepsilon 1}H$: 8.59; $C_{\delta 2}H$: 7.33
Gly-4	8.41	4.00-4.06		
Asp-5	8.37	4.63-4.71	2.66-2.83 [d]	
Gln-6	8.47	4.38	2.06 and 2.22	$C_\gamma H_2$: 2.42; $N_\varepsilon H_2$: 6.85 and 7.54
Gly-7	8.49	4.00-4.06		
Ser-8	8.24	4.45-4.52	3.83-3.97 [c]	
Asp-9	8.48	4.63-4.71	2.66-2.83 [d]	
Cys-10	8.32	4.53-4.59	2.98	
Ser-11	8.46	4.45-4.52	3.83-3.97 [c]	
Ile-12	8.01	4.20	1.91	$C_{\gamma 1}H_2$: 1.22 and 1.50; $C_{\gamma 2}H_3$: 0.96; $C_\delta H_3$: 0.90; CO-NH_2: 7.13 and 7.61
CH_3-CO				2.10

[a] $C_\alpha H_2$ group in the case of Gly residues; [b] $C_\beta H$ group in the case of the Ile residue; [c] Please note that the multiplets of the $C_\beta H_2$ resonances of all the three Ser residues fall in this range;
[d] Multiplets of the $C_\beta H_2$ resonances of the two Asp residues fall in this range

Table S2: Formation constants ($\log\beta$) of the proton and cadmium(II) complexes of **HS** (with the estimated errors in parentheses (last digit)) together with some derived data (I = 0.1 M $NaClO_4$, T = 298 K)

Species	pqr	$\log\beta_{pqr}$		$pK_{peptide}$
$[HL]^{3-}$	011	9.19(3)	$pK_{s,5}$	9.19
$[H_2L]^{2-}$	021	17.51(3)	$pK_{s,4}$	8.32
$[H_3L]^{-}$	031	24.04(4)	$pK_{s,3}$	6.53
H_4L	041	28.33(4)	$pK_{s,2}$	4.79
$[H_5L]^{+}$	051	31.83(5)	$pK_{s,1}$	3.50
				pK_{pqr} [a], $\log K_2^{L}$ [b], $\log K_2^{Cd}$ [c]
CdH_2L	121	22.67(8)	pK_{121}	4.43
$[CdHL]^{-}$	111	18.24(3)	pK_{111}	6.46
$[CdL]^{2-}$	101	11.78(4)	pK_{101}	10.36
$[CdH_{-1}L]^{3-}$	1–11	1.42(6)		
$[CdH_3L_2]^{3-}$	132	39.41(19)	pK_{132}	6.0
$[CdH_2L_2]^{4-}$	122	33.36(8)	pK_{122}	7.13
$[CdHL_2]^{5-}$	112	26.23(8)	pK_{122}	8.00
$[CdL_2]^{6-}$	102	18.23(7)	pK_{102}	10.88
$[CdH_{-1}L_2]^{7-}$	1–12	7.35(8)		
$[Cd_2HL]^{+}$	211	21.35(10)	pK_{211}	5.46
Cd_2L	201	15.89(7)	pK_{201}	9.27
$[Cd_2H_{-1}L]^{-}$	2–11	6.62(11)	pK_{2-11}	9.38
$[Cd_2H_{-2}L]^{2-}$	2–21	–2.76(6)	pK_{2-21}	11.15
$[Cd_2H_{-3}L]^{3-}$	2–31	–13.91(7)		
NP. [d]		610	$\log K_2^{L}$	6.45
FP (cm^3) [e]		0.003	$\log K_2^{Cd}$	4.11

[a] $pK_{pqr} = \log\beta_{pqr} - \log\beta_{p(q-1)r}$; [b] $\log K_2^{L} = \log\beta_{102} - \log\beta_{101}$; [c] $\log K_2^{Cd} = \log\beta_{201} - \log\beta_{101}$;
[d] NP = number of points; [e] FP = fitting parameter

Fig. S1.a-b: ^{1}H NMR spectrum of **HS** at pH = 5.37 with the assigned resonances ($c_{\mathbf{HS}}$ = 3.5×10^{-3} M, T = 298 K, H_2O/D_2O = 90:10 %v/v). Please note that the His-3 $C_\alpha H$ resonance is under the H_2O signal and its chemical shift was determined from 2D experiments, however, the chemical shift of the His-3 NH resonance could not be unambiguously determined.

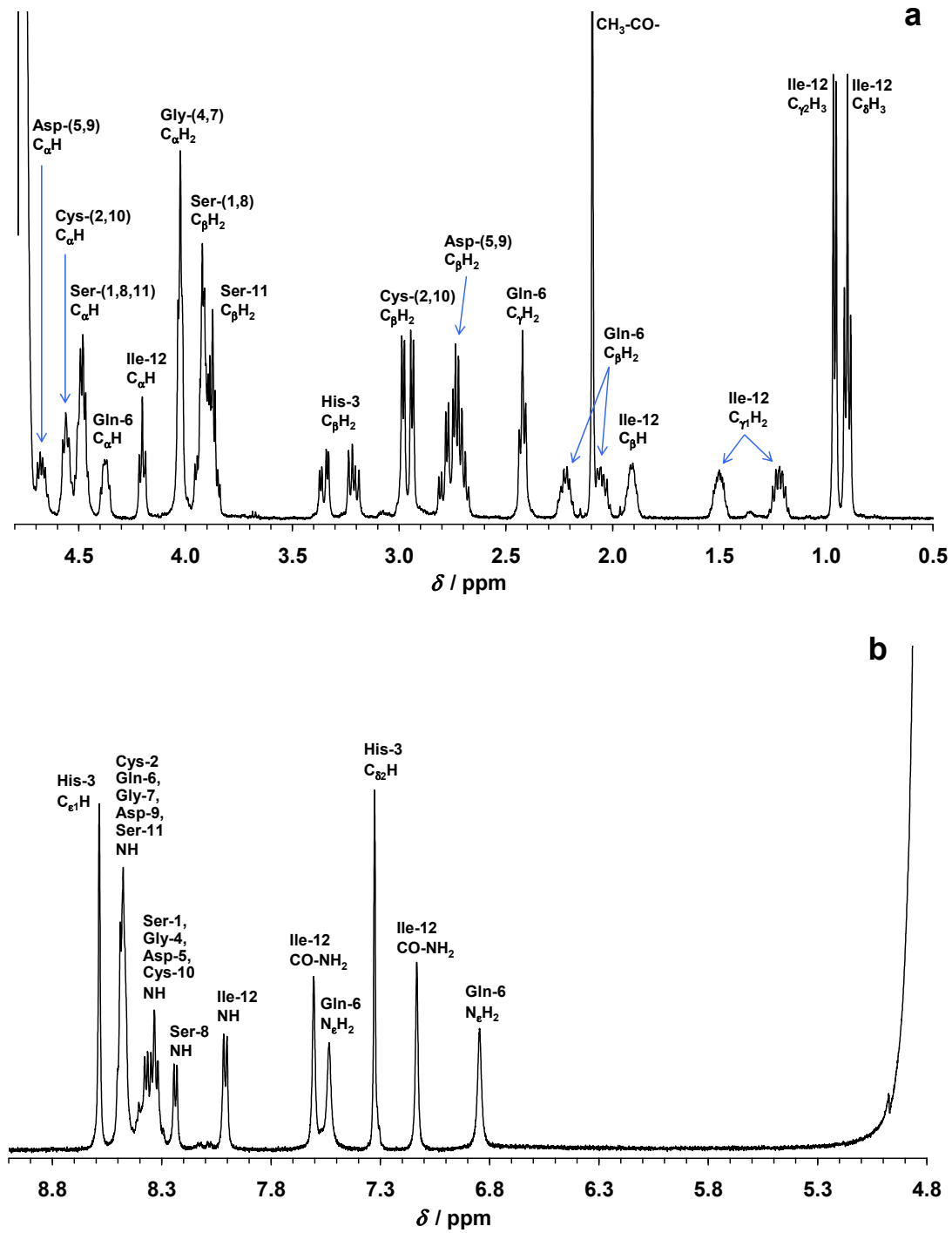
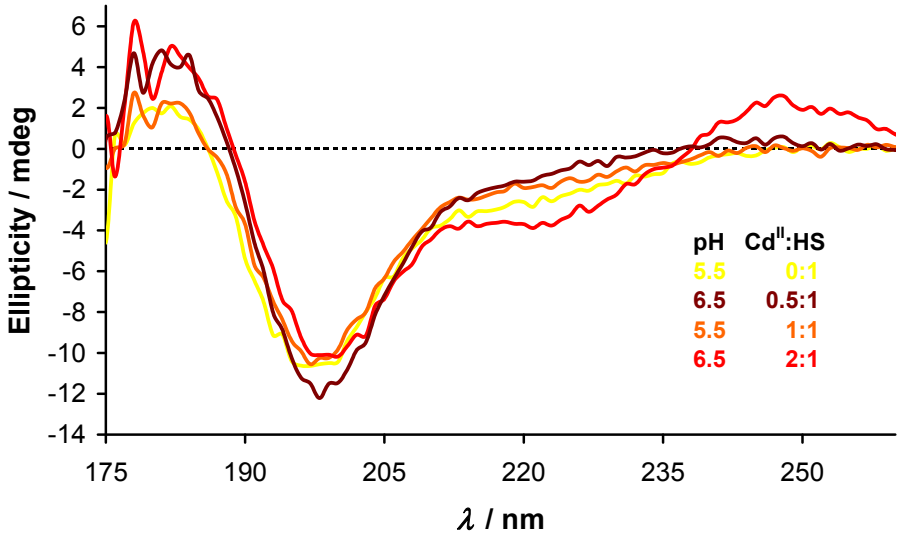


Fig. S2: The SRCD spectra of Cd^{II}-**HS** systems at different Cd^{II}:**HS** concentration ratios were recorded at the selected pH values to obtain information on the overall peptide conformation in different species.
(a) The spectra in the pH range of 5.5 - 6.5 show the difference between the spectral pattern of the CdHL with CdS_2 and CdH_2L_2 with CdS_2N_2 type of coordination. Also the special CD pattern for the system with metal ion excess is observed in the 210-260 nm range.
(b) The spectra in the pH range of 8.0 - 9.0 show the slight difference between the spectral pattern of the CdL with CdS_2NX and CdL_2 with CdS_4 type of coordination.

(a)



(b)

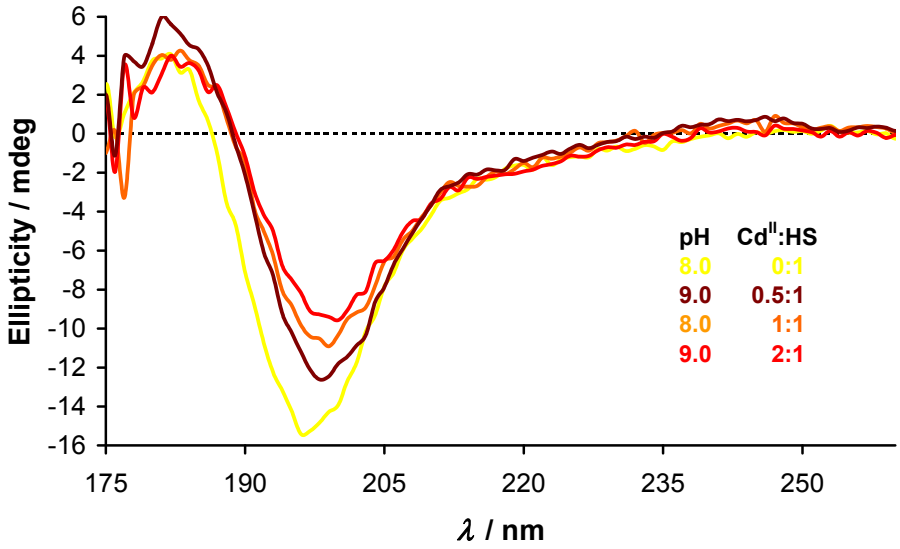
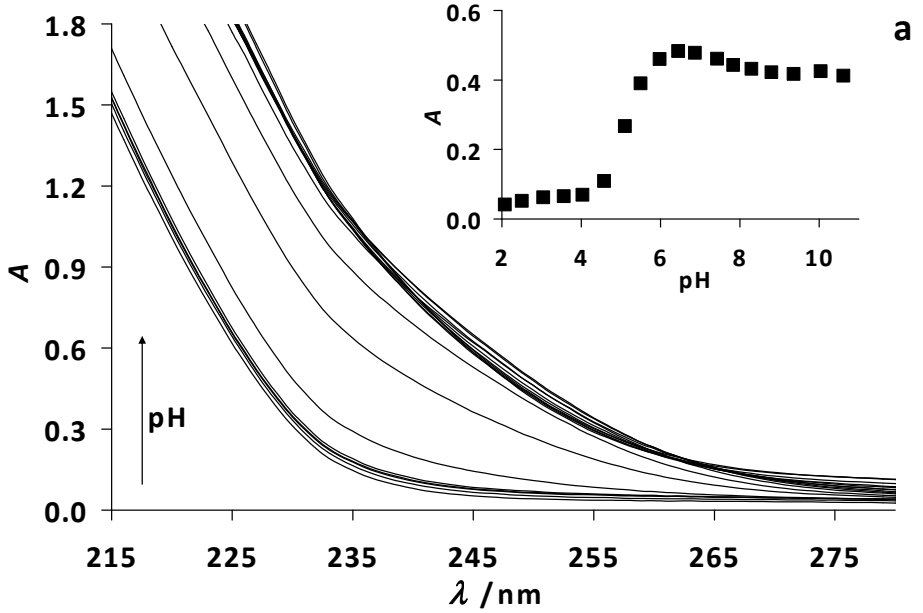
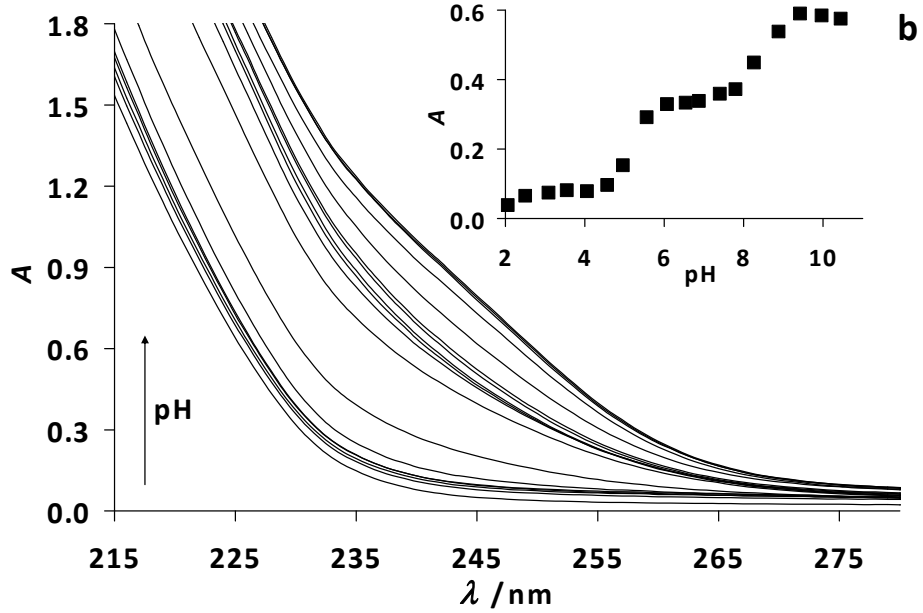


Fig. S3: pH-dependent UV spectra recorded in the cadmium(II) – **HS** 1:1 (a), 0.5:1 (b) and 2:1 (c) systems ($c_{\mathbf{HS}} = 1.0\times10^{-4}$ M (a,b,c), $I = 0.1$ M $NaClO_4$, $T = 298$ K). The inserts show the absorbance vs. pH traces at $\lambda = 250$ nm). Notice that the drop of intensity above pH ~ 8 in the presence of cadmium(II) excess (c) correlates with the deprotonation processes demonstrated by the speciation diagram (Fig. S7).





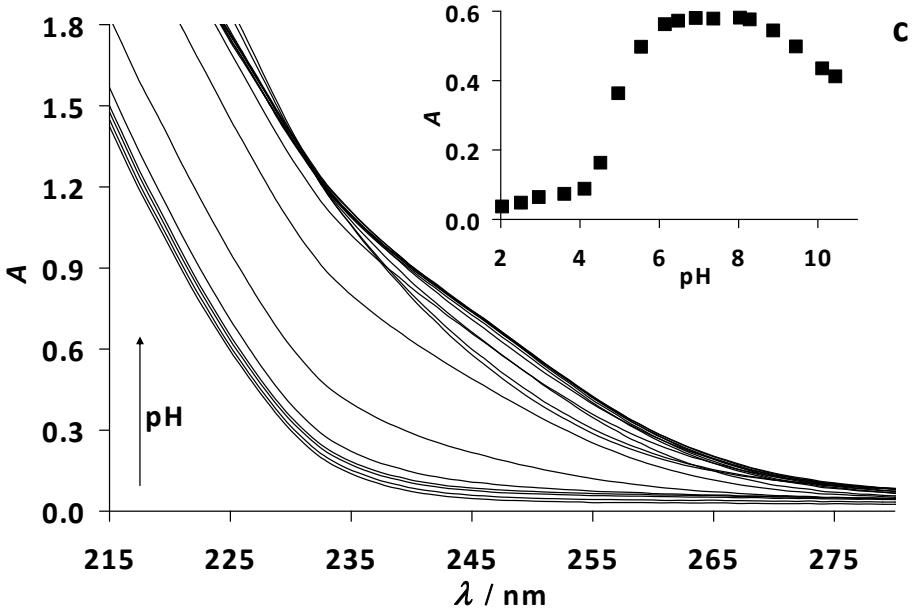


Fig. S4: pH-dependent ^{1}H NMR spectra of **HS** in the absence of metal ion. The symbols denote the following signals: His-3 $C_{\varepsilon 1}H$ (▲), His-3 $C_{\delta 2}H$ (△), His-3 $C_{\beta}H_2$ (◆),Cys-(2,10) $C_{\beta}H_2$ (✳) and Asp-(5,9) $C_{\beta}H_2$ (✕). ($c_{\mathbf{HS}}$ = 1.3×10^{-3} M, T = 298 K, H_2O/D_2O = 90:10 %v/v)

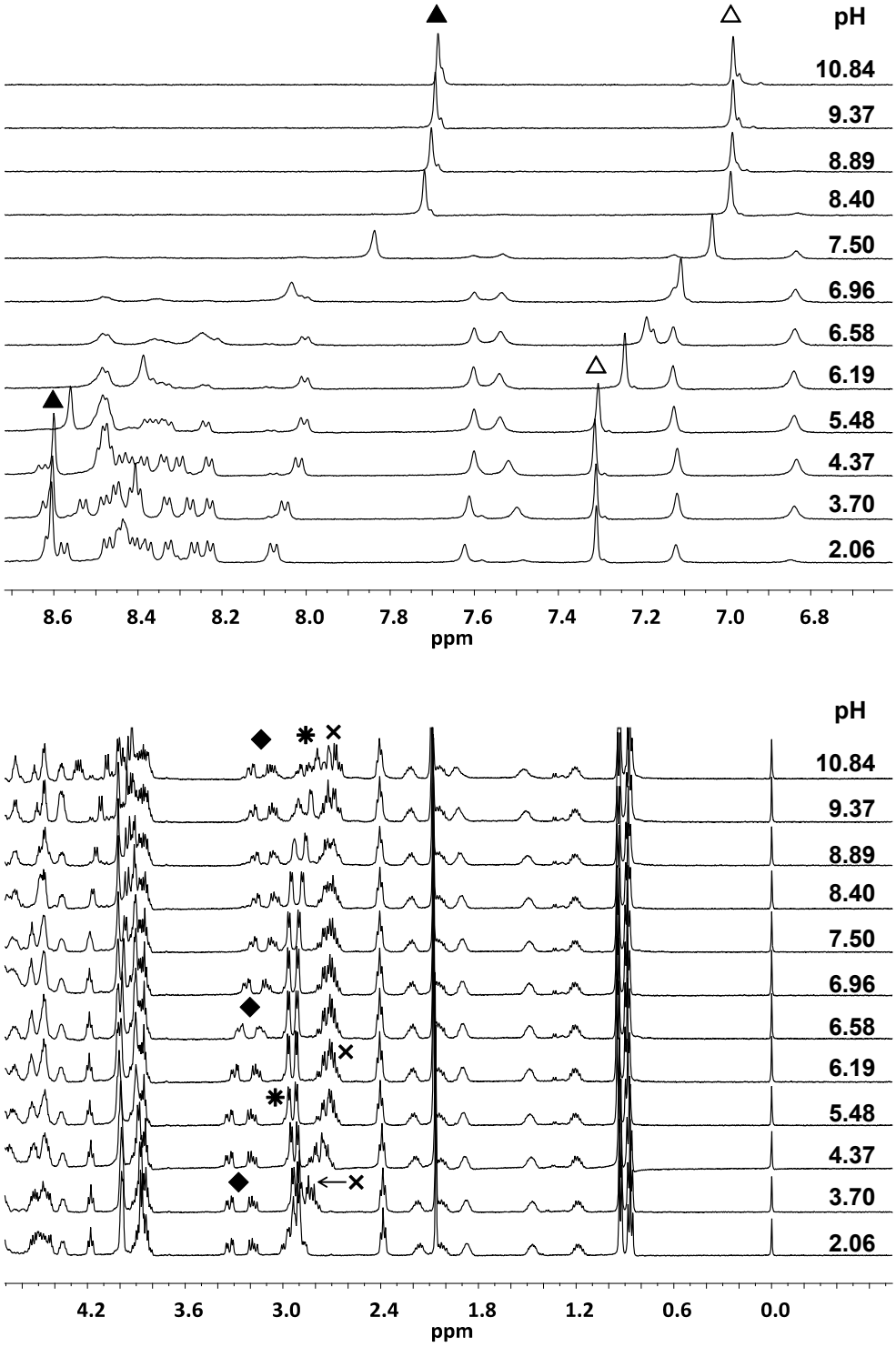
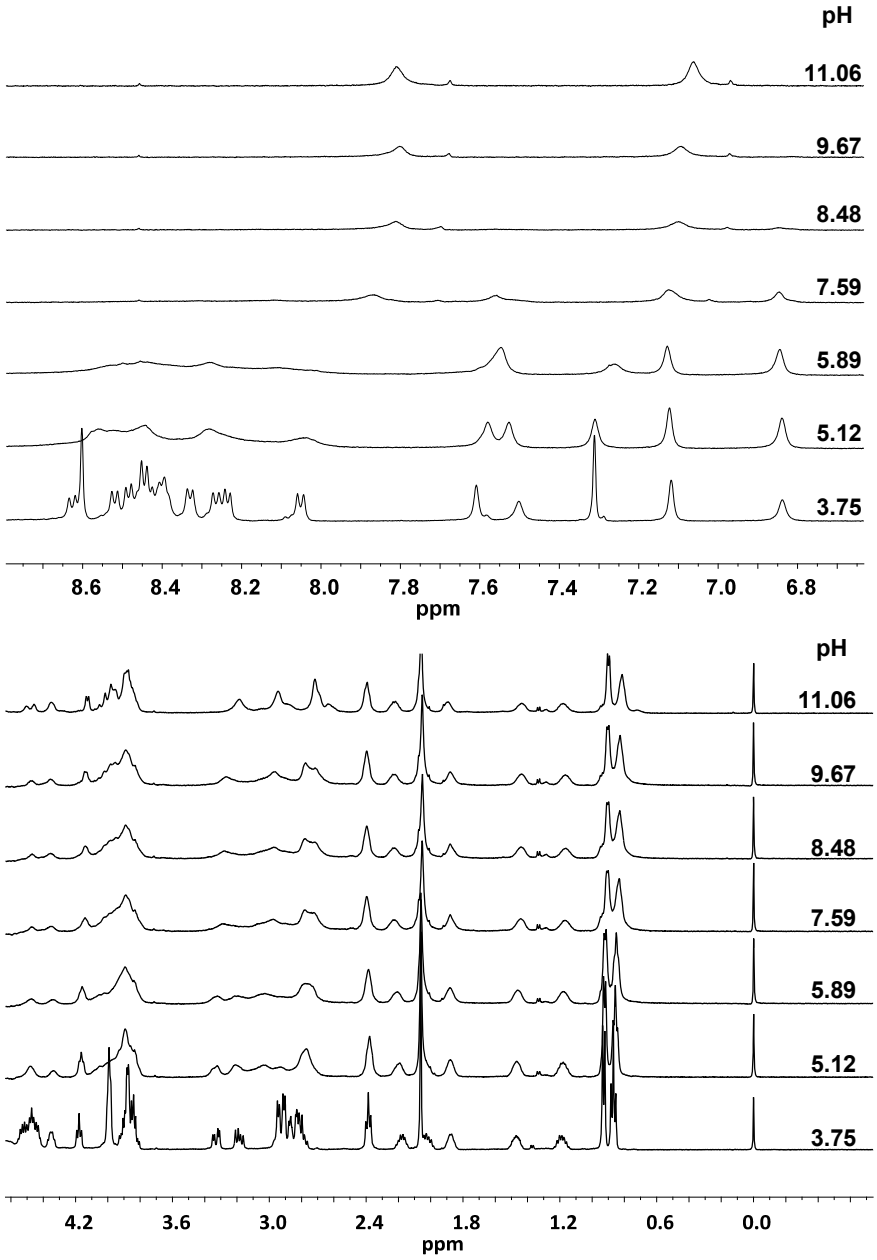


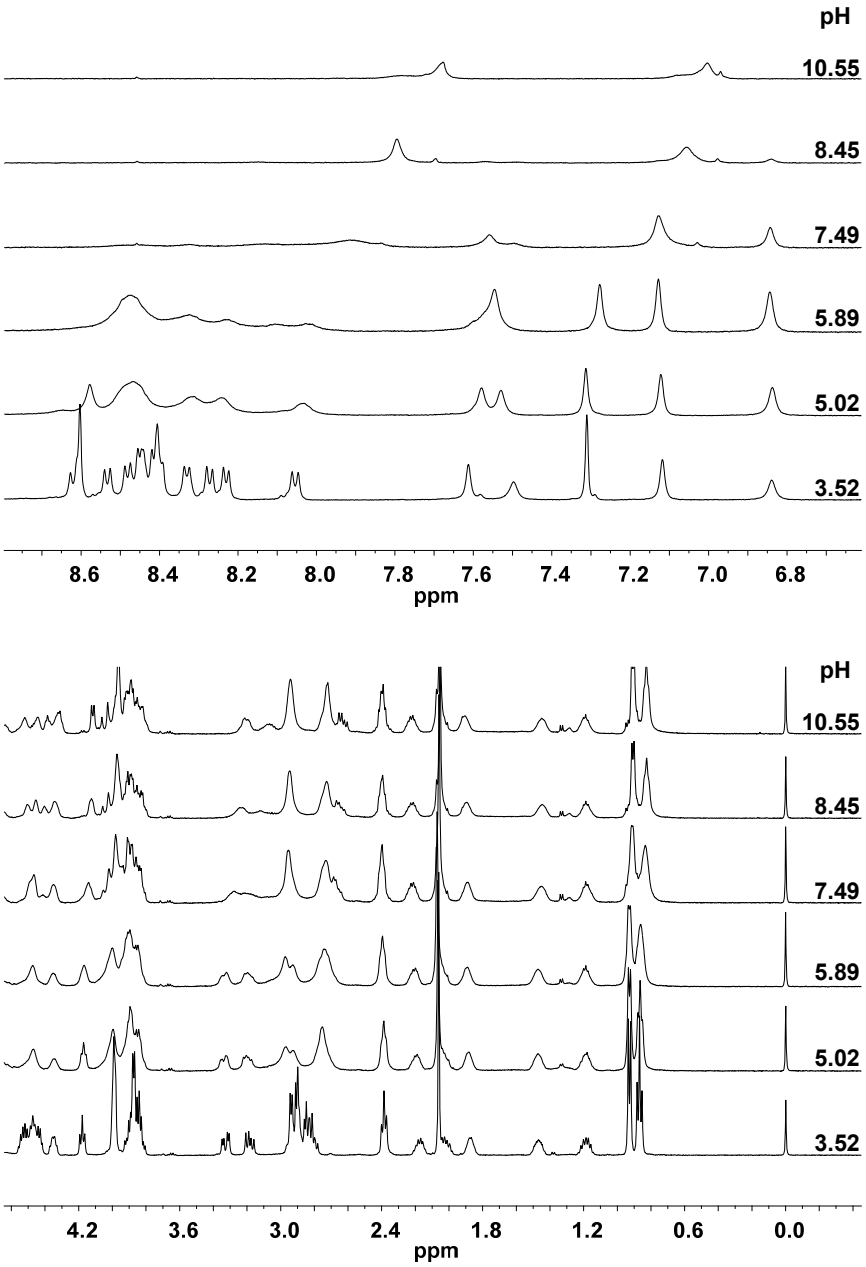
Fig. S5: pH-dependent ^{1}H NMR spectra recorded in the cadmium(II) – **HS** 1:1 (a), 0.5:1 (b) and 2:1 (c) systems ($c_{\mathbf{HS}} = 1.3\times10^{-3}$ M, $T = 298$ K, $H_2O/D_2O = 90:10$ %v/v)

c) Please note that a new set of signals appear in the presence of cadmium(II) excess as a consequence of the formation of the Cd_2L species, also reflected by the special shape of the SRCD spectra (Fig. S2) and the stronger intensity of the LMCT band around $\lambda = 250$ nm, as compared to the 1:1 Cd^{II}:**HS** system (compare the inserts of Fig. S3). According to the slow exchange with other complexes present in the solution, this species might be a dithiolate bridged dinuclear complex, existing in a wide pH range (Fig. S7).

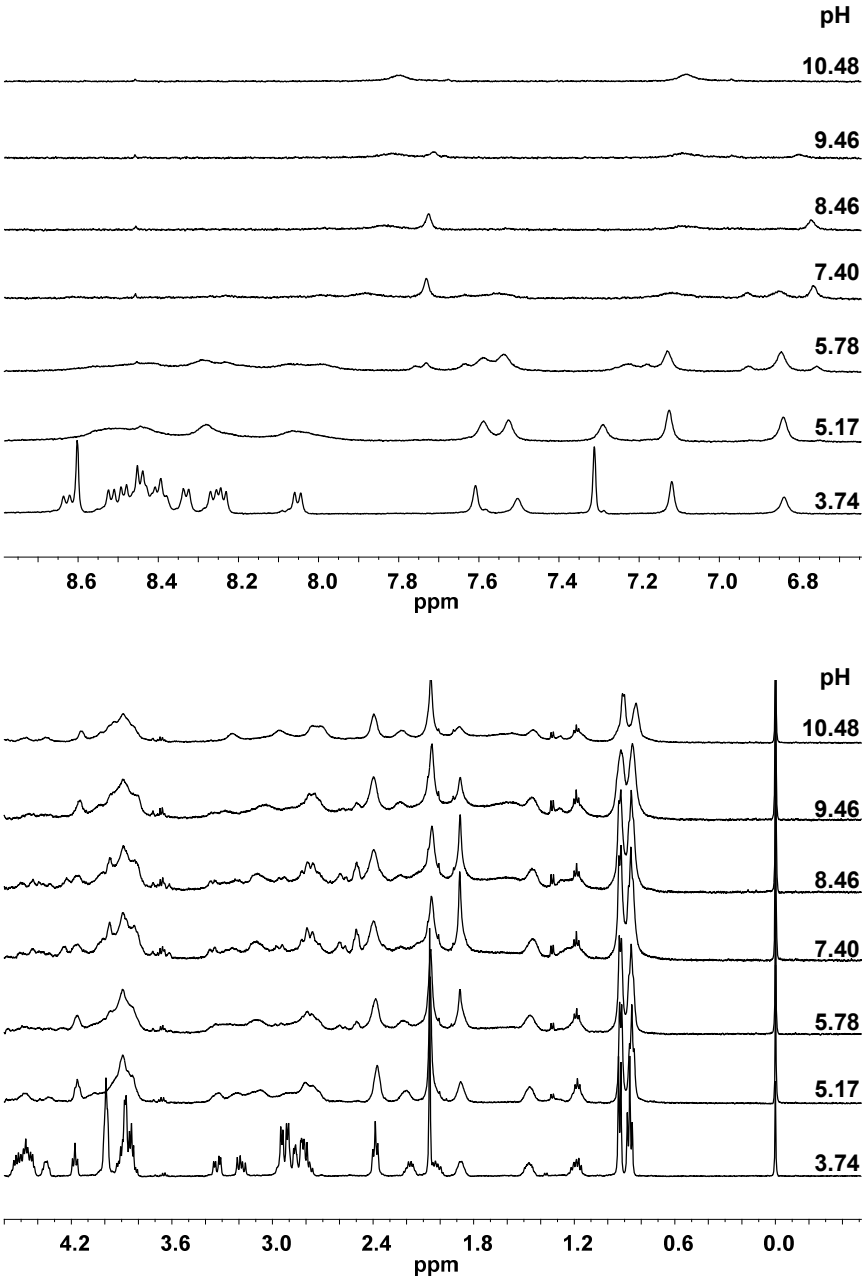
a)



b)



c)



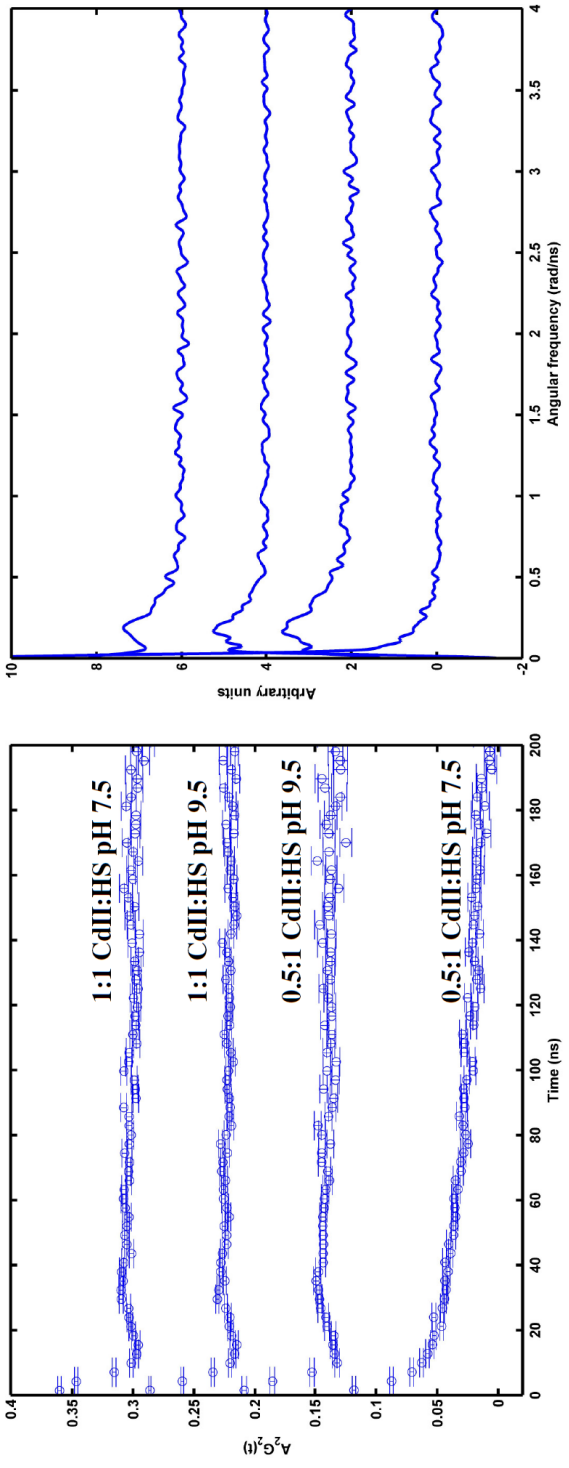


Fig. S6: ^{111m}Cd-PAC data. Left: Perturbation functions ; Right: Fourier transforms of the perturbation function.

Fig. S7: Species distribution diagram for cadmium(II) – **HS** 2:1 system (charges of the species are omitted for clarity). ($c_{\mathbf{HS}}$ = 1.0×10^{-3} M , I = 0.1 M $NaClO_4$, T = 298 K)

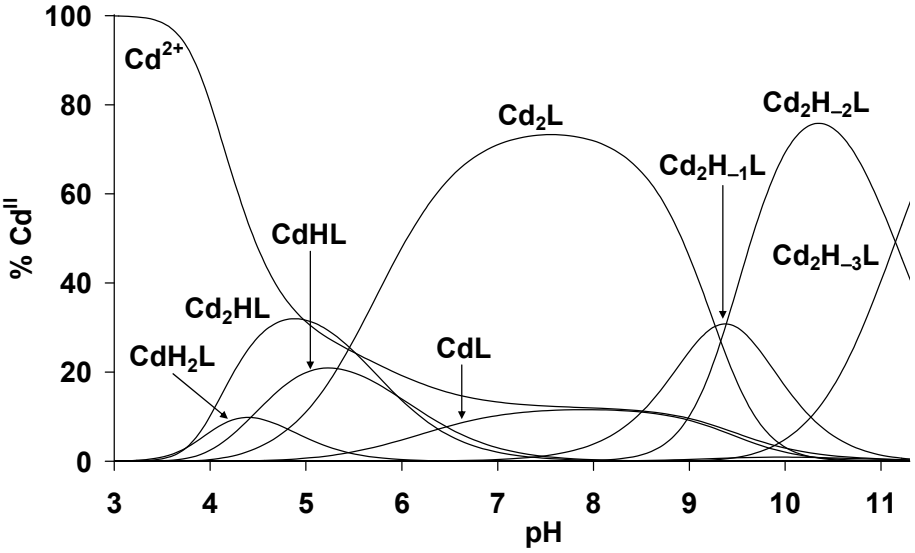
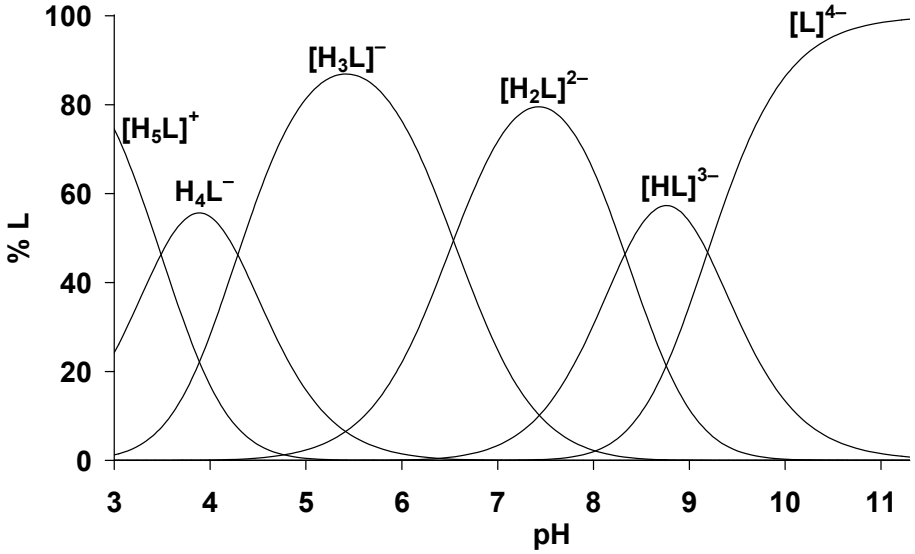


Fig. S8: Species distribution diagram for the peptide **HS** in the absence of metal ions ($c_{\mathbf{HS}}$ = 1.0×10^{-3} M (I = 0.1 M $NaClO_4$, T = 298 K).



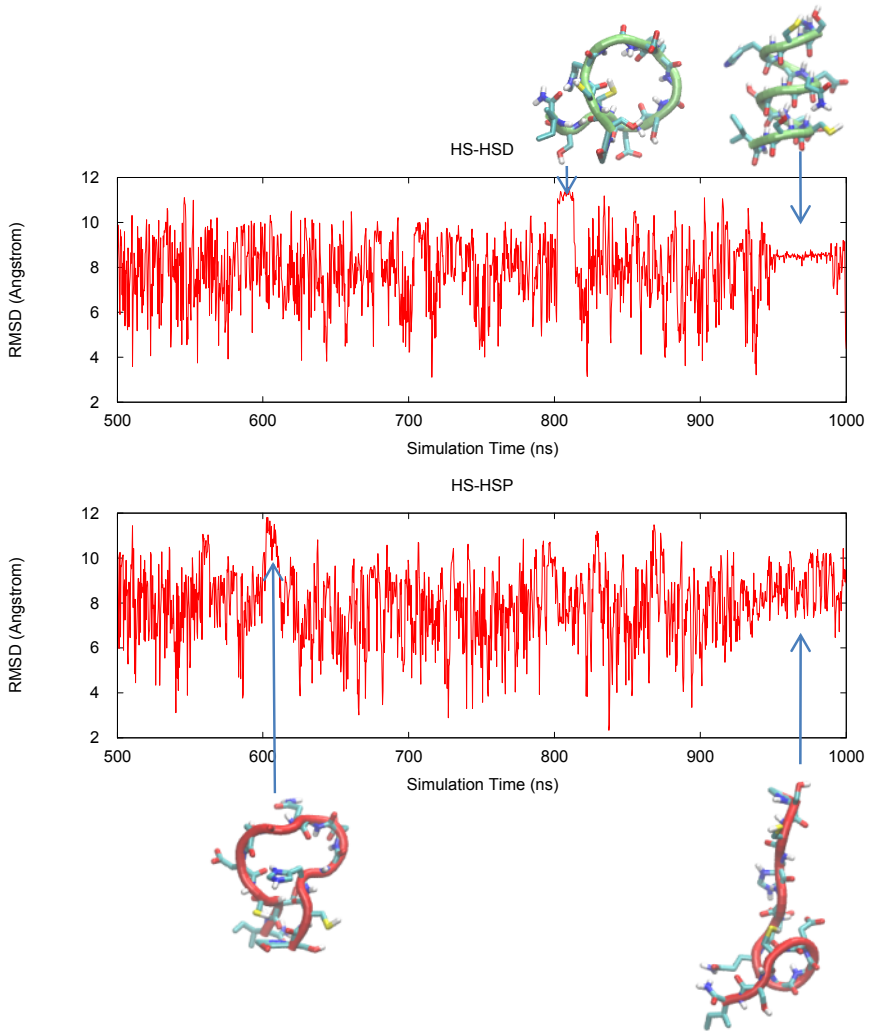


Fig. S9: Root mean-square distance (RMSD) versus simulation time for the HS-HSD (neutral His) and HS-HSP (protonated His) protonation states of the **HS** peptide. RMSD values are calculated based on backbone atoms with respect to the initial extended conformation of the HS-peptide. Selected structures corresponding to remarkable areas in the RMSD plots are shown. The complete set of conformations is contained in Movie S1.

Movie S1: Combined molecular dynamics trajectories from simulations of the HS-HSP (protonated His) and HS-HSD (neutral His) states of the **HS** peptide. The movie shows the last 500 ns of the 1000 ns simulation for each system. For reference, the pre-simulation extended structures of HS-HSD and HS-HSP have been inserted as the first frame of the movie. The frames are sampled at an interval of 400 ps. The backbones of HS-HSP and HS-HSD are color-coded in red and lime, respectively and all TIP3P waters as well as non-polar hydrogens have

been removed for clarity. It is clear from this movie and the RMSD plot in Figure S9 that a multitude of conformations with varying degrees of secondary structure content are visited during the simulation. The occupancy time for the different conformations exceeds tens of nanoseconds in some cases. Unique to the HS-HSD simulation is the population of a long-lived (~40 ns) all pi-helical state near the end of the simulation. A conformation characterized by a relative increase in secondary structure content is also found for HS-HSP at approximately the same point in the simulation, see Figure S9. The lifetimes of these states, however, are still fast on the timescale of NMR. This supports that fast exchange may cause the narrow resonance signals observed in ^{1}H NMR spectra for the free HS peptide.